

AN EVS CLICKER BASED HYBRID ASSESSMENT TO ENGAGE STUDENTS WITH MARKING CRITERIA

Steve Bennett, Trevor Barker and Mariana Lilley
University of Hertfordshire

ABSTRACT

Over 4 iterations of a large course (>180 students) in introductory emedia design in a first year computer science course we have seen a year on year improvement. We believe this is down to the use of evs clickers for feed-forward assessment: that is to say a method of getting the whole class to evaluate previous cohorts' submissions in public and discussing them, bringing to light the various properties they possess and how this maps to the marking rubric. This impacts on the students practices as they attempt their assignment. Over time, the practice has become more refined, principally through a rewritten criteria sheet, better training samples, and finally the development of a hybrid in-class assessment: the *swarmative* assessment combining both formative and summative practices and relying on its visibly social nature for its transformative power. This involves (a) evaluating previous submissions (in a non-graded way) – allowing for the free exercise of discriminative judgment not measured against any “authoritative” standard, but also (b) answering a set of objective questions about the work being assessed (what techniques were used to realize various effects). It ensures full cohort coverage together with engagement with the marking criteria.

KEYWORDS

Feed-Forward, EVS, Peer Assessment

1. INTRODUCTION

In the preceding 4 academic years, on a first year introductory Computer Science BSc module in multimedia development, the authors have used live feed-forward exercises using EVS clickers. Feed-forward essentially means getting students to mark previous students work to better understand the assignment criteria. Usually this is done in small groups with discussion. In our case, has involved using a lecture theatre with around 200 students together. In it we use two projectors one to display exemplars of previous students' work to be marked, and another projector to display graphs and average values of scores input by students in the audience using EVS clickers as they mark according to the rubric. For each previous assignment, students are led through the marking rubric and asked to mark the work as if they were the lecturer. Each dimension of the work being assessed has 5 attainment descriptors – students have to choose a number from one to five. When everyone has voted, then the average score is shown to the room as well as a histogram of numbers of students voting for each attainment level.

What they are evaluating is a Multimedia CV assignment. This is a vehicle for students to be tested on their ability to design an interface and to be able to manipulate media to ensure the smallest possible file size. This is to test a learning outcome of learning how computers digitize and store media and the various filesize reduction techniques available. Students are asked to create 5 screens in their CV: a home page with an animation on it, and four other pages for work, hobbies, education and contact details. It is therefore a certainly small assignment, but one which does require some planning, some sense of structure (particularly for the implementation of the audio player which needed to be independent of the other navigational actions of the user, i.e moving between screens) a house style, very small file size, and finally designed in such a way as to permit rapid resizing in the one hour final adjustments section before submission.

This technique (of getting students to mark previous “exemplar” work) has been done before, but not in live public sessions. Notable examples are O'Donovan, Price and Rust [2004], Hendry, Armstrong and Bromberger [2011] and Wimshurst and Manning [2012]. The theoretical basis for this kind of procedure comes most clearly from D Royce Sadler [1987] who beautifully expresses how just a few concrete

instantiations of seeing the marking criteria used in practice can make the whole experience of understanding and purposefully striving to achieve learning outcomes so much more comprehensible for students.

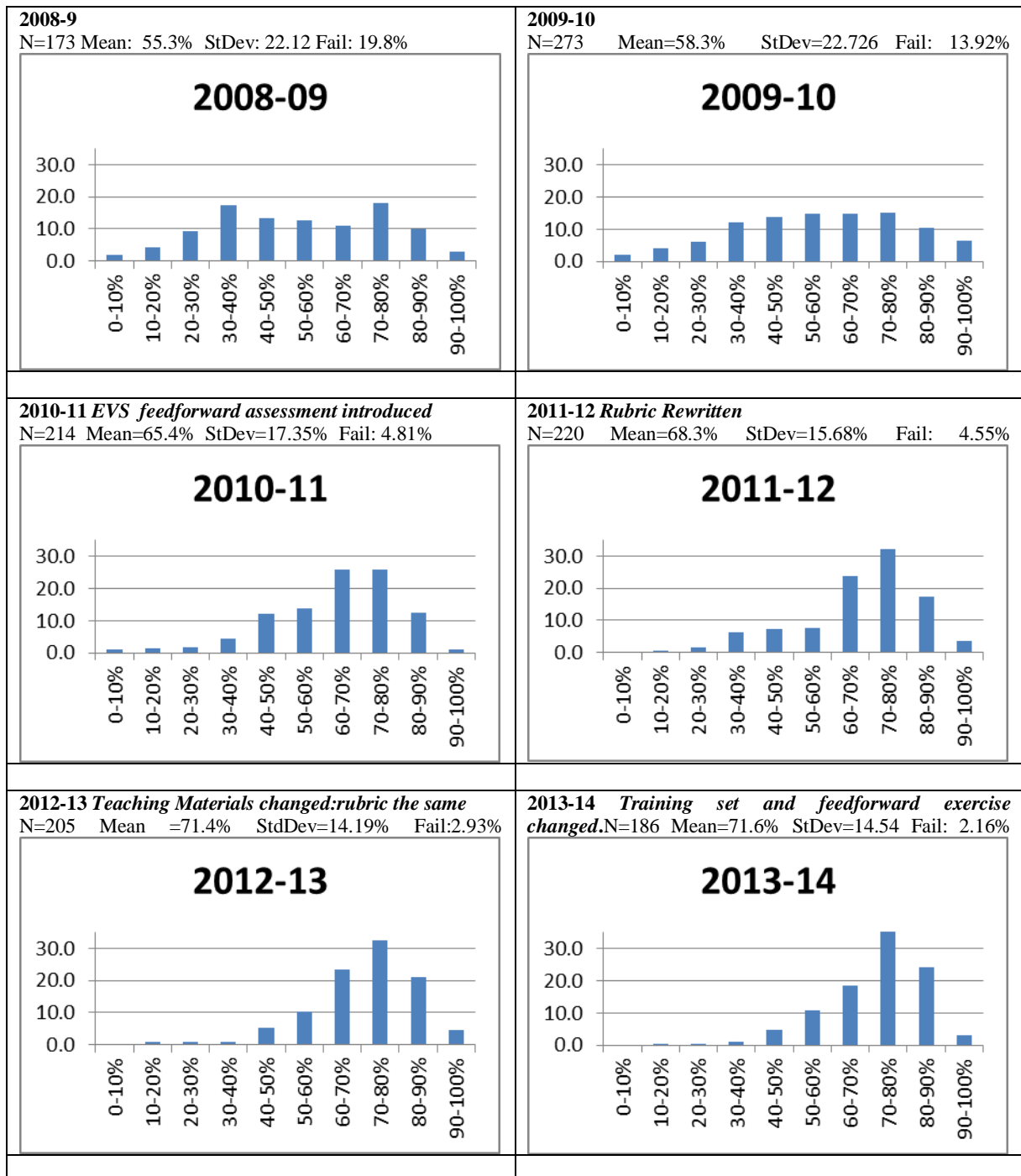
This area of research has much in common with the more researched area of student peer-assessment (see particularly the meta studies conducted by Topping [1998], Falchikov and Goldfinch[2000], Zundert et al[2010] and much of the claimed benefits are in common (particularly in the area of meta-cognition). However, feed-forward, because it does not impact on the actual grades of the student whose work is being assessed (because they did it for a previous cohort) it may be more effective at improving meta-cognition in the assessor. The major practical difference between the work described in the papers above and ours is its public nature and immediacy.

2. LIVE FEEDFORWARD AND STUDENTS' REACTIONS TO IT

In order to achieve the engagement of the entire cohort, each year we run this evaluation event twice: the first time formatively - looking at work done by previous cohorts, marking them and discuss the marking. The second time it is run summatively (credit bearing) – where students received a mark according to the level of agreement between their grading and that of the tutors (though this is marked extremely generously). However, this caused hostility in the first and third year it was used. Therefore in the last occasion we created hybrid assessment type – which we shall call the *swarmative* assessment – involving a formative part evaluating previous students work (which was not graded) followed by a small 10 question test on some of the practical features and techniques used in those previous artifacts which could be marked on an objective basis. It is there both summative and formative, however we argue that the power of this technique of live-feedforward occurs in the way it causes the student to compare the kinds of marks they are awarding with those given by the rest of the class.

2.1 E-Media Design Historically

The six below represent the scores for a large final assignment on the course. The x axis represents the final student percentage score in bands of 10%. The y axis represents the percent of the cohort in each decile. In this list, only students scoring above zero have been included. Although the first intervention using live feedforward occurs in 2010-11, we have included the two previous years in order to demonstrate that the results obtained in 2009-10 were not merely the result of the vastly larger cohort in that particular year.



As can be seen, a significant improvement takes place not only in the indicators of average grade, but also in terms of the shape of the distribution. In 2009-10, the distribution becomes more normal, and shows a small overall improvement in average grade. 2010-11 was the first time we used the feed-forward method and did so with the identical criteria used in the 2009-10 assessment. This small improvement however came at the cost of some student hostility owing to the manner in which the feed-forward was conducted: namely, students were given marks by how near their marking pattern coincided with that of the tutor. Students complained of being asked to guess what we the tutors were thinking.

In 2011-12 therefore we rewrote the criteria to remove the subjectivism from the more basic things being tested (for instance “has the required number of screens”) and used likert scales for the more higher order criteria (appropriateness of design for instance). In 2012-13 the core course material for learning Adobe Flash was changed since the previous material had become outdated, though the assignment and rubric remained the same.

In 2013-14 with the rubric only marginally changed, the modality of undertaking the feed-forward was changed. Just as in previous cases there would be a practice session and then there would be a marked second session – however, the marked second session involved both a formative assessment of the works being evaluated, as well as a small summative test on the *objective properties of those works* (for instance what technique allowed a certain effect to appear). Additionally, in response to the criticism of the quality of the training set, we introduced a new much better training set – producing past CVs but anonymizing the students’ personal details, but also putting pictures of another person of the same age, gender and ethnicity in their stead. This was meant to answer the concerns raised both about the training set and also about the method used for crediting participation in the feed-forward events.

Though there are a number of variables in play, it does seem that quality of the overall student work has increased over that time, as well as the fact that it seems to have advantaged the entire cohort, and we believe that this is in large part due to the use of the feed-forward techniques. At this point, the question becomes: what is it that students are doing differently in their work to achieve this improvement?

The principal evidence we will look at will be from two focus groups undertaken with the 2012-13 (**Em2012**) cohort and the 2013-14 (**Em2013**) cohort where students from the course were asked to discuss their experience on it. Some secondary informing evidence will also be used from focus groups used on other courses: a humanities course on Web Publishing from 2012-13 (**Hum2012**), a joint focus group of the two 2012-13 courses together as well as one technology and engineering course (**EmAHum2012**), and a focus group from a 2011-12 Masters Course in Multimedia Computing (**MmSpec2011**). While the modalities of full cohort evaluation sessions were different on the other courses (Humanities involved a small class size with live evaluation primarily being used as a prompt for discussion) and the Multimedia masters course being used for full peer evaluation, nonetheless some of the feelings expressed by participants in those courses are generalizable to the participants in E-Media Design.

3. REGULATING STUDENT EFFORT

On the premise that students achieving higher grades are either doing more work or doing more efficient and goal focused work we sought to find out how students regulated their effort of the course of an assignment. This course is also interesting since it takes place in the first semester of their university career when they are making the transition to university study.

In this light, how do students get a sense of what they are aiming for? Both focus groups mentioned the evaluation session as being crucial to getting a sense of the level required. The primary benefit was the clarity of the criteria. One student said:

“definitely appreciated the seeing all the CVs - actually that was something I thought of when we were talking about the differences with going to university...saying I want x y and z, also there’s a proven example of what the finished product might look like, and what you thought of these different finished products that you’ve currently got up - an assignment where it’s like, it seems like it’s clear what we are supposed to do, but all of us are going, I’ve no idea, and we don’t see a finished product, so seeing the [examples], was really good.” (**Em2012**)

This backs up what Royce Sadler has said about the value of concretizing abstract attainment criteria in specific exemplars which embody them. But while his research has typically focused on the understanding of higher-order criteria, the benefits also help with the satisfaction of the lower order criteria. Particularly, it taught students what to avoid

“I think we could have all just looked at the CVs and gone that is a good one, that is a bad one but when you guys went through it and explained why that is a good one and why that is a bad one and pointed out the bad things and the good things then everyone got a better idea of what you expected of us. So like you said, the basic problems didn’t come up because you guys told us what the basic problems were so we all avoided them” (**Em2012**)

In 2012 one student said

“I took like things from it and I saw things not what to do there were some things like maybe the sound was not appropriate or - like the buttons and the colour contrast that just clashes I had to guidance from that but it was good” (Em2012)

As well as being able to avoid simple errors – it also helped with impasse management. Having a clear sense of what they are trying to achieve, seemed to help them identify points at which they were unable to get to their goal – and how to get round it (usually by recourse to help from fellow students).

A student from the EM2013 focus group said:

“My friend asked me if I could help her on the sound because they couldn’t get the sound buttons to pick up the music and I had forgotten how to make music like package sorter.” (Em2013)

In the previous year, another student took a long term approach to the development of her project:

“I looked at the Mark criteria and I thought, I am going to do what I can and if there are any extra bits I can do. I will try and do it then maybe I can try and ask for help. Because didn’t we have the holiday during that period or something and when we come back from the holiday then I will ask at see if anyone can help me with the extra bits that I find difficult.” (Em2012)

In other cases, students mentioned collaborating with fellows or with partners. In the 2013-14 group one student mentioned a facebook group of approximately 4 people in which he participated – in the 2012-13 group a student talked of working together with others in the LRC (Learning Resources Centre) for bouncing off ideas about the most appropriate sound to use. He said

“the most complicated bit was adding the sound. The CV itself didn’t take that long about creating it everything was all right. But then me and my friends were at the LRC and we were all talking about how to put it in which one to use I think that was the worst thing because I would suggest one sound and then put it in the CV and it wouldn’t correspond with what I was talking about.” (Em2012)

“Appropriateness” a term only used in the higher order criteria seemed to generate the most reflection. Another student from that year said

“And choosing sound it took quite a while because I find something and then I’d question whether was really appropriate or not so then I’d have to look again and it was quite confusing what would be seen as appropriate” (Em2012)

Another student talked about the major benefit of the marking exercises as being:

“making sure and got all the buttons uniform all the same size all the text inside the buttons all the same size and all appropriate” (Em2012)

Therefore the marking rubric seems to have done its job both at the lower end (causing students to avoid obvious pitfalls) but also at the upper end (causing sustained reflection, both individually and in groups about appropriateness and suitability). Also the clarity of the desired objective (obtained from seeing the exemplar works) has helped students metacognitively reflect on their means for achieving the goal.

However, the rubric has only gone through two iterations – and the improvement of marks was already in evidence before the rewriting of the current rubric. Therefore what seems to have happened is that the ritual of the collective marking activity seems to have enhanced the *authority* and *tangibility* of the rubric. From the 2009 date we can see there was a rubric – but it certainly wasn’t attended to by the students – regardless of the detailed description of how marks were arrived at, students seemed to have not attended to it. However, in the more recent iterations of the course, the rubric has come to be the corner stone around which their work revolved. One particularly pragmatic student said

(Student)The sheet that you get basically gave us a list and markings of what you guys expected so it kind of told us exactly what to do.

(Tutor) If it told you exactly what to do, while you were writing out your CV you were kind of ticking bits off?

(Student)Well what I did was I read through all of it, then did my CV, then read through it all again and checked everything and compared it with my CV and then changed things around it.

(Tutor) That is really interesting. Did you two do the same?

(Student)Yes, pretty much. I just followed the requirements sheet.

(Tutor) And did you do it formally as in ticking things off or did you do it...

(Student)I used the (00.07.07) and come back to them eventually but I tend to do the easiest bit or the bits that I am not struggling with.

(Tutor) So you did the easiest bits and sort of ticked them off and then you come back to the harder bit later. What were the harder bits?

(Student)The ones that involved more – making a background colour in grey is easy it just involves doing something in stuff but making things move around or something that involves (00.07.45) so you would come back to that. (Em2013)

Here we clearly see the centrality of the rubric, its role in structuring work-flow and the attention to the easier and planning for the more complex, and in the earlier quotations, we saw how it also governed the occasions when students sought help from others. All of this is deeply similar to the *participation-reification* duality described in Wenger's Community of Practice[1987]. The rubric represents a *reification* and codification of the practices necessary for a successful multimedia CV. The *participation* is the kind of work students do outside lectures and practices which make that successful multimedia CV – but also comprises the informal collaborations, help seeking and work planning which we have noticed above. However, as we have seen before, the mere fact of a rubric does not mean students will take notice of it, or structure their academic activity around it. But the faith in it, its authority and centrality comes from the ritual of the EVS public evaluation. Having learnt how to apply the rubric, by collectively marking 3-6 pieces of work together, the core centrality of the rubric becomes apparent to the students.

4. THE SWARMATIVE ASSESSMENT

A problem that has endured since first attempting this intervention is how to ensure compulsory participation in it. Were we to make participation a purely voluntary experience the danger is that the benefits would only extend to the attendees. In the first 3 iterations we ensured this by actually grading the students' evaluation of previous exemplars – that was done by measuring the difference between the scores the tutors awarded and the scores the students awarded. In a focus group, one student said:

I think the exercise is very good because it kind of gave you an idea of what was expected what you were looking for. The assessment of trying to determine what you would have marked them on, came across a bit harsh. The fact that we were being assessed on, speculating on, on what we think you had marked it on. So it was good in the sense that it gave you an idea of what to watch out for when you are doing your own...but because you were trying to assess on what someone else thought about somebody else's work, it was like a double stage hop (Em2012)

In order to resolve this problem – requiring compulsory participation but without grading according to marking co-incidence we designed a hybrid assessment. This required the students to attend a session – the first part of which would involve the evaluations of a previous cohort's artifacts which was not grade bearing. However, the second half would involve answering 10 objective questions about those artifacts (which tool was used to achieve effect x) which would be grade bearing – this meant it required everyone to be present – but the grade bearing part was a truly objective test and based on coincidence of subjective judgment.

We justify the use of the word “swarm” because powerful effects can happen when students mark in public. They can compare the mark they awarded with the marks given by the cohort. One student said:

for example you see websites and you and to you it's great and you confidently put excellent and when the rating comes out everybody says it was rubbish and wondering why did I? how did that come about? And so it challenges your intellectual you know thinking (Hum2012)

5. CONCLUSION

Chickering and Gamson [1987] choose “communicates high expectations;” as one of the seven principles for good practice in undergraduate education. The use of exemplars has been reported in a number of studies as being an effective way to achieve this. However, this typically takes place in small group situation involving discussion between tutors and students and is labor and resource intensive. We believe the technique we have shown here, offers a way of efficiently transmitting these expectations a very large cohort. Having an evaluation event involving both the subjective evaluation of exemplars as well as some objective test on the properties or techniques used in those exemplars appears the best way to satisfy the goal of full cohort engagement with assignment marking criteria.

REFERENCES

- O'Donovan, B, Price, M, Rust C, 2004, "Know what I mean? Enhancing student understanding of assessment standards and criteria" *In Teaching in Higher Education*, vol. 9, no. 3, pp. 325-335,
- Hendry, Graham D., Nikki Bromberger and Susan Armstrong, 2011, "Constructive guidance and feedback for learning: The usefulness of exemplars, marking sheets and different types of feedback in a first year law subject" *In Assessment & Evaluation in Higher Education* 36, vol. 1, pp. 1-11,
- Wimshurst, Kerry and Matthew Manning, 2013, "Feed-forward assessment, exemplars and peer marking: evidence of efficacy" *In Assessment & Evaluation in Higher Education* 38.4, pp. 451-465,
- Sadler, D. Royce, 1987. "Specifying and promulgating achievement standards", *Oxford Review of Education* 13.2, pp. 191-209,
- Topping Keith, 1998, "Peer assessment between students in colleges and universities", *Review of Educational Research* 68.3, pp. 249-276,
- Falchikov, Nancy and Goldfinch.Judy, 2000, "Student peer assessment in higher education: A meta-analysis comparing peer and teacher marks", *Review of educational research* 70.3, pp. 287-322,
- Van Zundert, Marjo, Sluijsmans Dominique and Van Merriënboer Jeroen, 2010, "Effective peer assessment processes: Research findings and future directions", *Learning and Instruction* 20.4, pp. 270-279,
- Chickering, Arthur W., and Zelda F. Gamson, 1987, "Seven principles for good practice in undergraduate education." *AAHE bulletin* 3 (7): 7,
- Wenger, Etienne, 1998, *Communities of practice: Learning, meaning, and identity*. Cambridge university press.